AL/HR-TP-1995-0039

ARMSTRONG LABORATORY

MODELING PRACTICE, PERFORMANCE, AND LEARNING

Valerie J. Shute
J. Wesley Regian

HUMAN RESOURCES DIRECTORATE
TECHNICAL TRAINING RESEARCH DIVISION
7909 Lindbergh Drive
Brooks AFB, Texas 78235-5352

Lisa A. Gawlick

Galaxy Scientific Corporation
1880 Carswell Avenue
Lackland AFB TX 78236-5507

January 1996

Interim Technical Paper for Period October 1994 - October 1995

AIR FORCE MATERIEL COMMAND
BROOKS AIR FORCE BASE, TEXAS

NOTICE

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto:

The Office of Public Affairs has reviewed this paper, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

Valerie J. Shute

VALERIE J. SHUTE
Project Scientist

R. Bruce Gould

R. BRUCE GOULD
Technical Director
Technical Training Research Division

James B. Bushman

JAMES B. BUSHMAN, Lt Col, USAF
Chief, Technical Training Research Division
Human Resources Directorate

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	January 1996	Interim Technical Paper - October 1994-October 1995

4. TITLE AND SUBTITLE

Modeling Practice, Performance, and Learning

5. FUNDING NUMBERS

C - F41622-92-D-0006
PE - 61102F
PR - 2313
TA - T2
WU - 18

6. AUTHOR(S)

Valerie J. Shute   Lisa A. Gawlick
J. Wesley Regian

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Galaxy Scientific Corporation
1880 Carswell Avenue
Lackland AFB TX 78236-5507

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Armstrong Laboratory (AFMC)
Human Resources Directorate
Technical Training Research Division
7909 Lindbergh Drive
Brooks AFB TX 78235-5352

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

AL/HR-TP-1995-0039

11. SUPPLEMENTARY NOTES

Armstrong Laboratory Technical Monitor: Dr. Valerie Shute (210-671-2734)

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release: distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This chapter presents the results from a study examining the relationship between practice, performance, and learning. We compared two versions of an intelligent tutoring system differing only in the number of problems that needed to be solved per problem set (Abbreviated = 3 problems, Extended = 12 problems). Our hypotheses were that Abbreviated subjects, in comparison to Extended subjects, would: (a) take less time to complete the tutor because they had fewer problems to solve, (b) perform worse on the posttest measures (accuracy and latency), and (c) demonstrate poorer transfer of knowledge and skills across tutor problems given fewer practice opportunities. We found that, while Abbreviated subjects did take significantly less time to complete the tutor than Extended subjects, both groups performed equally across all outcome measures. Componential skill analyses enabled us to track the course of skill acquisition during practice, and predict the degree of skill transfer afterward. We conclude with suggestions for the development of efficient automated instruction.

14. SUBJECT TERMS

Cognitive abilities, componential skill analysis, intelligent tutoring systems, modeling, learning (outcome and efficiency), performance, practice effects, transfer (lateral and vertical)

15. NUMBER OF PAGES

22

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

PREFACE.

The research reported in this paper was conducted by personnel at the Armstrong Laboratory, Human Resources Directorate, Brooks Air Force Base, Texas. The opinions expressed are those of the authors and do not necessarily reflect those of the Air Force.

Correspondence concerning this paper should be addressed to Valerie J. Shute, Armstrong Laboratory (AL/HRTI), 1880 Carswell Avenue, Lackland Air Force Base, Texas, 78236-5507.

Modeling Practice, Performance, and Learning

SUMMARY

This paper presents the results from a study examining the relationship between practice, performance, and learning. We compared two versions of an intelligent tutoring system differing only in the number of problems that needed to be solved per problem set (Abbreviated = 3 problems, Extended = 12 problems). Our hypotheses were that Abbreviated subjects, in comparison to Extended subjects, would: (a) take less time to complete the tutor because they had fewer problems to solve, (b) perform worse on the posttest measures (accuracy and latency), and (c) demonstrate poorer transfer of knowledge and skills across tutor problems given fewer practice opportunities. We found that, while Abbreviated subjects did take significantly less time to complete the tutor than Extended subjects, both groups performed *equally* across all outcome measures. Componential skill analyses enabled us to track the course of skill acquisition during practice, and predict the degree of skill transfer afterward. We conclude with suggestions for the development of efficient automated instruction.

Introduction

The purpose of this paper is to examine the effects of practice on within-tutor performance and learning outcome. The relationship between practice and performance addresses the issue of how practice influences learning rates, errors, and the degree of successful transfer *during* the learning process (Singley & Anderson, 1986). And the relationship between practice and outcome addresses how practice affects what learners ultimately walk away with at the *end* of a learning task, including retention, application, and transfer to some novel task. Both of these relationships are believed to follow the well-documented tenet: "Practice makes perfect (Ackerman, 1988; Anderson, 1987; Bryan & Harter, 1899; Fisk & Rogers, 1992; Schneider & Shiffrin, 1977; Woltz, 1988) as well as the related tenet: When the number or variety of example problems is restricted, skill acquisition tends to be rapid, but transfer tends to be weak (Carlson & Yaure, 1990; Gick & Holyoak, 1987). In this paper, we submit both convictions to an experimental test to determine just how much practice makes perfect, what is weak transfer, and so on.

Another relationship we are interested in exploring, but which is not so clear, exists between performance and outcome. It seems reasonable to assume that acquisition performance is a good indicator of "learning." However, Schmidt and Bjork (1992) have shown how, relative to a "standard" condition, practice environments that show little improvement, or even *decreased* performance during skill acquisition, may actually produce increased outcome performance. What is ultimately learned may therefore be poorly reflected by characteristics of acquisition. However, one way to examine the performance-outcome relationship is in terms of transfer.

Transfer of skills can pertain to either learning performance or outcome data. One can transfer skills, problem-to-problem, during the learning process, or one can transfer skills from the conclusion of some learning task to a novel task. Gagné & Paradise (1961) have distinguished these two kinds of transfer (i.e., vertical and lateral) through a framework of skill hierarchies. *Vertical* transfer involves transfer from lower- to higher-level skills that exist in part-whole, prerequisite relationships to one another (e.g., transfer of learning between

problem sets within a tutor where the curriculum has been hierarchically arranged). *Lateral* transfer deals with transferring abilities from one situation to another (e.g., transfer of learning from tutor- to posttest environment). If prerequisite skills are attained, then acquisition of more complex knowledge and skills would be expected. One may find, however, that even though the low-level skills have been successfully acquired, some learners may fail when they reach the complex skills. This problem can be corrected by presenting additional practice on the high-level skills (Smith, 1986).

To illustrate how practice may differentially impact performance and outcome, we describe the results from a large-scale experiment using an intelligent tutoring system (ITS) teaching flight engineering skills. This study was designed to test practice effects using an ITS originally developed at the University of Pittsburgh (Lesgold, Bunzo, McGinnis, & Eastman, 1989) then systematically altered at the Armstrong Laboratory to fit experimental objectives (Shute, 1993). Job components included collecting and analyzing information about a pending flight and deciding whether various factors (e.g., weather and runway conditions, type of plane) indicate a safe flight. There were two parts to the tutor's curriculum: (a) The Graph section, teaching the basic (component) knowledge and skills used by a flight engineer (e.g., reading and interpreting graphs), and (b) The TOLD (Take Off and Landing Data) Worksheet, requiring an integration of skills learned in the first section of the tutor.[1] The tutor consisted of 23 problem sets: 14 in the Graph section and 9 in the TOLD section. This paper focuses on the learning data from the Graph section of the tutor because it constituted the instructional portion of the tutor and produced more of a modeling challenge. That is, the Graph section contained a wide range of new knowledge and skills spanning 14 problem sets. A listing of each problem set, labeled G-1 to G-14, is presented in Appendix 1.

The tutor was manipulated to yield contrasting practice conditions, differing only in the number of problems the learner needed to solve in each of the problem sets. The version of the tutor consisting of many problems was called "Extended" (12 problems per problem set). And the version with fewer problems was called "Abbreviated" (3 problems per set). We believed that this ratio (4:1) provided a reasonable contrast between practice conditions. Twelve problems correspond to a typical number of problems presented in exercise sections of textbooks, and constitute a large enough number to examine learning curves. The solution of only three problems was believed to be a minimal number to support the acquisition of novel concepts and skills.

The usual hypothesis from experiments varying practice schedules is that the manipulation will impact learning efficiency and/or outcome (i.e., large between-group differences in rate of skill acquisition or degree of attained skill). Thus, we hypothesized that subjects assigned to the Abbreviated condition would take less time to complete the tutor because there were considerably fewer problems for them to solve. However, these same subjects were not expected to perform as well on the posttest measures (accuracy and latency) compared with subjects learning from the Extended condition, who would have received considerably more practice in the problem-solving environment (lateral transfer). In addition, given fewer problem-solving opportunities, Abbreviated subjects were expected to show relatively weak vertical transfer of knowledge and skills across successive problem sets during tutor-learning in comparison to the Extended subjects.

[1] A detailed explanation of the individual problem sets for both the Graph and TOLD sections of the tutor can be found in Shute (1993).

Findings from the Flight Engineering Study

The effects of the two practice conditions were investigated in relation to learning time (hours needed to complete the tutor's curriculum), learning outcome (posttest latency and accuracy measures) and parameters of skill acquisition (errors and problem solving times during tutor learning). Approximately 370 subjects participated in the study, randomly assigned to one of the two practice conditions -- Extended or Abbreviated. All subjects were obtained from a temporary employment agency, and paid for completing the study. None of the subjects had any formal training or experience with the subject matter instructed by the tutor.

The first hypothesis was supported: Subjects in the Abbreviated condition required significantly less time to complete the tutor compared to subjects in the Extended condition (about 7 versus 12 hr, respectively, for total time on tutor; 4 versus 6.5 hr for Graph-section times). Notice that while the Abbreviated subjects received 1/4 the problems as the Extended subjects, they required more than 1/2 the amount of time to complete the tutor. Thus, the Abbreviated subjects spent relatively more time per problem than the Extended group. This will be examined in more detail in the section of this paper on "Performance Differences Between Conditions."

The second hypothesis tested whether subjects in the Extended condition would perform better on the posttest than subjects in the Abbreviated condition, given their additional problem-solving opportunities. This hypothesis was not supported: Subjects in the Abbreviated and Extended conditions performed *equally* on both percent correct and latency data for all outcome measures. Table 1 shows the accuracy and latency measures, per group.

Table 1
Differences in Outcome Measures by Practice Condition

TEST/MEASURE	ABBREVIATED	EXTENDED	SIGNIF
Percent Correct			
Declarative Knowledge	61.4 (23.8)	61.0 (24.1)	NS
Procedural Knowledge	46.9 (22.6)	48.6 (24.2)	NS
Procedural Skill	62.8 (19.3)	65.7 (19.4)	NS
Latency (mean sec. per item)			
Declarative Knowledge	15.3 (06.3)	14.4 (06.5)	NS
Procedural Knowledge	25.6 (12.2)	24.3 (09.6)	NS
Procedural Skill	20.8 (07.8)	20.1 (08.1)	NS

Note. Means are presented with standard deviations in parentheses.

This unexpected finding prompted us to conduct a more fine-grained analysis of the data. First, we sought to determine whether these findings were due to the posttest being an insensitive measure of knowledge and skill acquisition (e.g., restricted range, ceiling or floor effects). Next, we decomposed the tutor's problem sets into component skills, and examined subjects' performance data in relation to those low-level skills. Finally, the component skills were used as building blocks for modeling performance and learning data. In the following sections, we describe these steps.

Quality of the Posttest

The first issue concerns the psychometric properties of the posttest (i.e., reliability and validity measures). The posttest was designed to measure the acquisition of tutor-specific

(flight engineering) knowledge and skills, and was divided into three parts: Declarative knowledge, Procedural knowledge, and Graph reading and interpretation of various charts (i.e., Cartesian coordinate, Polar coordinate, Maximum Allowable Crosswind, and Wind Components -- a Polar coordinate chart superimposed on a Cartesian coordinate chart). Declarative knowledge items related to definitions of relevant concepts learned from the tutor (e.g., What is a headwind? The relative wind direction is a function of what two components?). Procedural knowledge questions required subjects to apply different rules, in their proper sequence (e.g., If the relative wind direction is 100 degrees, what is its wind type and normalization value?). For the graph interpretation section, subjects were shown specific graphs that were used during the tutor, and were required to solve problems relating to those graphs. The posttest was made up of 46 items, and was administered on the computer in a multiple choice format, with six alternatives to choose from. Chance performance was thus 17% for each of the three parts.

We conducted an item analysis on the posttest data and found neither ceiling nor floor effects, or any group differences for any of the items (note: the data were analyzed altogether, as well as separately, by condition). The items turned out to be good discriminators of acquired knowledge and skill (i.e., none were too difficult or too easy); each item showed about a 50% accuracy level. We also found the test to be a *reliable* measure: odd-even reliability = .85, overall. To test the *validity* of the posttest, we compared specific error-type data from the tutor (e.g., errors dealing with incorrect use of the Wind Components Chart) with a comparable categorization of posttest items (e.g., items dealing with the use of the Wind Components Chart). Correlations between the tutor-error and outcome data were high (Graph-section r = .70; TOLD section r = .74). Finally, we examined whether subjects actually learned anything from the tutor. Chance performance on the posttest = 17%, and the posttest data were normally distributed with a mean of 58% (SD = 21). Although we did not collect pretest data from the current group of subjects, we did collect pretest data in another study using the same tutor and subjects from the same population of temporary service employees. In that study, pretest Mean = 38% (SD = 11.5; N = 90) and posttest mean = 58%, the same as this population's mean. So, most subjects did appear to learn from the tutor, compared to chance performance as well as to a similar group's pretest -- posttest data.

We next examine within-tutor data to see whether the Abbreviated subjects performed differently from the Extended subjects during skill acquisition to possibly explain the outcome results.

Performance Differences Between Conditions: Latency and Accuracy Data

We summed subjects' latency data as well as their accuracy data across the first three problems for each of the 14 problem sets. The first three problems were ones that both groups had to solve, so this sum represented performance data on mutual problems (i.e., Abbreviated subjects only solved three problems, while the Extended subjects solved three, then an additional nine problems, per problem set). Items were randomly administered to subjects; thus for the Extended subjects, the first three problems may not have been identical to the problems given to the Abbreviated subjects. These data are shown in Figure 1 (latency and accuracy data.

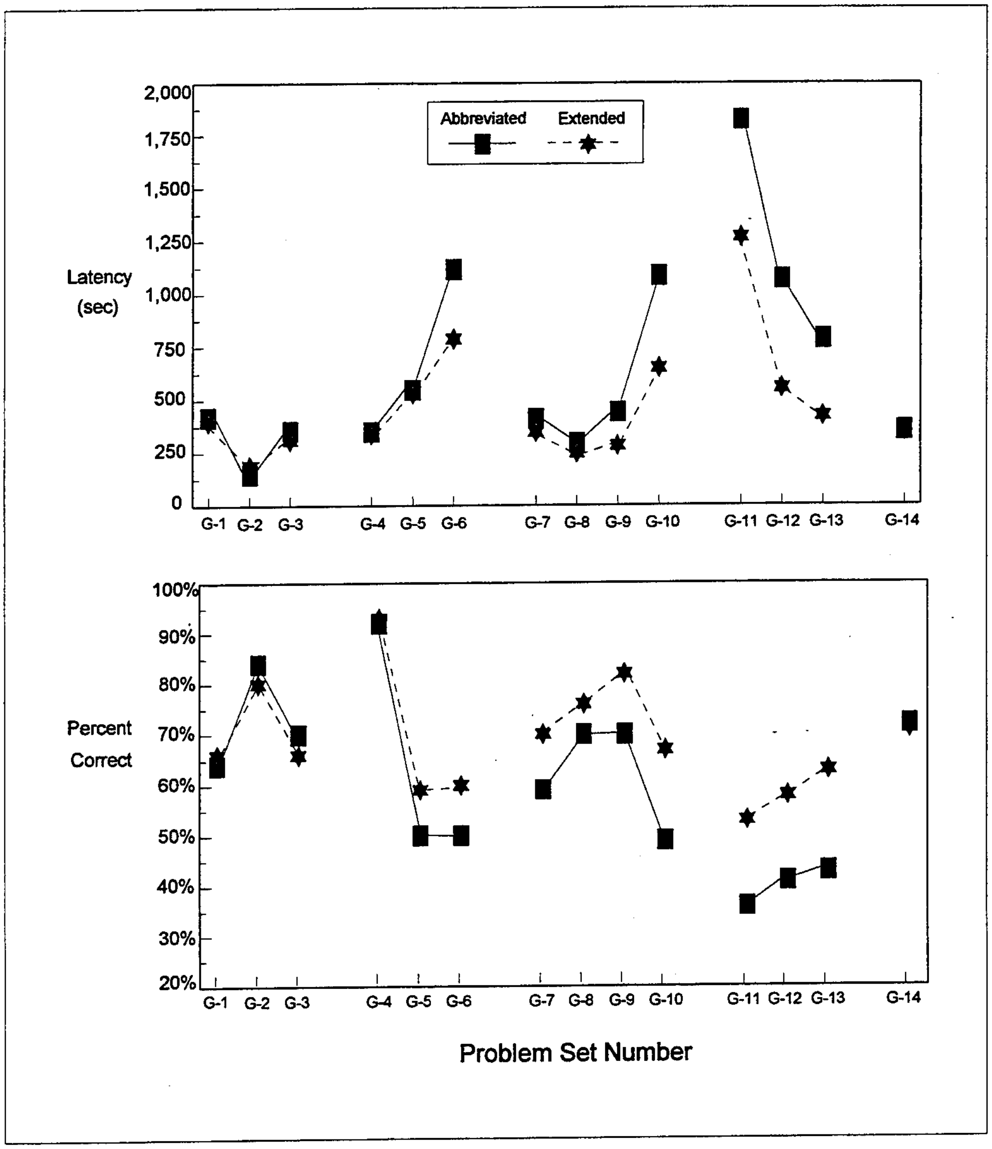

Figure 1

Latency and Accuracy Data from Problem Sets in Graph section of the Tutor

Latency data from the tutor showed that, for the more difficult problem sets (i.e., G-6, G-9 through G-13), Abbreviated subjects required more time on their three problems compared to the Extended subjects. But was there any evidence of a speed-accuracy tradeoff? That is, were the Abbreviated subjects more accurate on problems in which they had taken more time? In both conditions, subjects needed to answer a problem correctly before progressing to the next problem within any given problem set. For example, in the Abbreviated condition (with only three problems to solve per problem set), a subject could answer problem 1 incorrectly, then try again and answer it correctly. Similarly, the subject could answer problem 2 incorrectly then correctly. And finally, the subject could answer problem 3 correctly on the first trial. That would translate into a "percent correct" score of 60% for that problem set: 3 correct out of 5 attempts. The same scoring scheme was used for subjects in the Extended condition. The answer to the speed-accuracy tradeoff issue was no: wherever significant latency differences appeared between groups, subjects in the Abbreviated condition were also less accurate on these problem sets compared to the Extended subjects.

We then investigated performance data in terms of relative degree of *transfer* demonstrated per group, across problem sets. Learning curves were compared and the findings were as follows. First, both groups demonstrated roughly parallel learning curves across problem sets, where latencies and accuracies predictably increased or decreased (shown in Figure 1). That is, when a problem set introduced a lot of *new* knowledge and skills (e.g., G-10), latencies and errors increased dramatically, per group (albeit, somewhat more dramatically for the Abbreviated group). Similarly, when a problem set involved knowledge and skills that had already been acquired, subjects' data in both groups showed decreases in latency and increases in accuracy (e.g., G-13). These findings suggest a comparable degree of vertical transfer for both groups. Next, we attempt to model these data to see just how predictable they are.

Componential Skill Analysis and Modeling of Data

Appendix 1 lists the 14 problem sets of the Graph section of the tutor, arranged into four categories based on similar knowledge and skills: (a) Conversion Scales (G-1, G-2, G-3), (b) Cartesian Coordinate Grids (G-4, G-5, G-6,G-14), (c) Polar Coordinate Charts (G-7, G-8, G-9, G-10), and (d) Wind Components Charts (G-11, G-12, G-13). Within each category, problem sets increased in difficulty due to: stricter mastery criteria, inclusion of new knowledge and skills (both quantity and difficulty levels), or whether the problem represented an analog to previous problem sets in the same cluster. In this section, we decompose problem sets into their component skills, and model the within-tutor performance data as a function of skill acquisition status (i.e., classification of component skills as being learned, not learned, or unknown). We also attempt to model posttest performance based on tutor performance data (i.e., predict how an individual will perform on individual posttest items given the items were classified the same as tutor problem sets).

Componential Skill Analysis. Before making specific predictions of within-tutor performance and learning outcome data, we analyzed the degree to which problem-set clusters contained overlapping skills. High correlations among component skills within the same cluster could explain the vertical transfer evidenced by both groups, as well as validate the tutor's rationally-designed curriculum. Intercorrelations were computed on the latency data, and the mean correlations by problem-set cluster were: Conversion Scale Problems: 0.35**; Cartesian Coordinate Grid: 0.47**; Polar Coordinate Chart: 0.42**; and Computing Wind Components: 0.40** (** $N = 364$; $p < .001$). The accuracy data showed a similar pattern and magnitude of correlations. Thus, these correlations suggest that there are overlapping skills among problem-set clusters.

To confirm the existence of shared skills, we decomposed each problem set into its component skills, defined in terms of new, as well as previously introduced skills. This listing can be seen in Appendix 1. Notice that quite a few problem sets, especially within the same cluster, contain the same skills (e.g., Apply vertical and horizontal rulers in the Cartesian Coordinate Grid cluster). Altogether, we established 25 unique skills across problem sets (see bottom section of Table 2, "List of Unique Skills"), and there were 9 skills, on average, per problem set (note: earlier problem sets had fewer skills compared to later, more difficult problem sets).

Table 2
Matrix of Skills and Accuracy Data for Example Subject (R.J.)

	Scale Conversions			*Cartesian Coordinate*				*Polar Coordinate*				*Wind Components*		
	G-1	G-2	G-3	G-4	G-5	G-6	G-14	G-7	G-8	G-9	G-10	G-11	G-12	G-13
PC (%)	100	100	100	100	100	14	75	75	100	100	75	8	22	29
Skill														
1	1	1	1	0	0	0	0	0	0	0	0	0	0	0
2	1	1	1	0	0	0	0	0	0	0	0	0	0	0
3	1	1	1	1	1	0	1	0	1	1	1	1	1	1
4	1	1	1	0	1	1	1	1	0	1	1	1	1	1
5	0	0	1	0	0	1	0	0	0	0	0	0	0	0
6	0	0	0	1	1	0	1	0	0	0	0	1	1	1
7	0	0	0	1	1	1	1	0	0	0	0	0	1	0
8	0	0	0	1	1	1	1	0	0	0	0	1	0	1
9	0	0	0	1	0	1	0	0	1	1	1	1	1	1
10	0	0	0	0	1	0	1	0	0	0	0	0	0	0
11	0	0	0	0	0	1	0	0	0	0	0	0	0	0
12	0	0	0	0	0	1	0	0	0	0	0	0	0	0
13	0	0	0	0	0	1	0	1	0	1	1	1	1	1
14	0	0	0	0	0	0	0	1	0	1	0	1	1	1
15	0	0	0	0	0	0	0	1	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	1	1	1	0	0	0
17	0	0	0	0	0	0	0	0	1	1	1	1	1	1
18	0	0	0	0	0	0	0	0	1	1	1	1	1	1
19	0	0	0	0	0	0	0	0	0	0	1	1	1	1
20	0	0	0	0	0	0	0	0	0	0	1	1	1	1
21	0	0	0	0	0	0	0	0	0	0	1	1	1	1
22	0	0	0	0	0	0	0	0	0	0	0	1	1	1
23	0	0	0	0	0	0	0	0	0	0	0	1	1	1
24	0	0	0	0	0	0	0	0	0	0	0	1	1	1
25	0	0	0	0	0	0	1	0	0	0	0	0	0	0

List of Unique Skills

1. Find given value on conversion scale
2. Read to other side for equivalent value
3. Interpolate/determine value
4. Input value
5. Concept: altimeter setting
6. Cartesian coordinate chart
7. Use vertical ruler
8. Use horizontal ruler
9. Plot point at intersection
10. Straight-line graph
11. Correction table (correcting pressure altitude)
12. Separate altimeter setting into x/y coordinates
13. Compute value
14. Relative wind direction (RWD)
15. Determine wind type (via rules)
16. Polar coordinate chart (PCC)
17. Plot vector
18. Locate arc
19. Quartered, symmetric PCC (folded in 1/2, 1/4)
20. Normalize RWD
21. Draw new vector on chart
22. Wind components chart (Cartesian + Polar)
23. Gust element (add to TW & XW vel., not HW)
24. Extraction location (HW/TW - y, XW - x axis)
25. Locate given RCR on chart

Table 2, above, displays an example of one subject's matrix -- 25 skills by 14 problem sets. Matrices were created for each subject, containing a series of 1's and 0's denoting whether a skill was present (required) or absent (not required) for that particular problem set. For example, problem set G-1 only required skills 1 - 4. Therefore, values in the G-1 column

of this matrix contained 1's for skill 1 to skill 4, while the remainder of the cells contained 0's. We also included problem-set accuracy (percent correct) data in each matrix (row 1).

Three "status categories" of skills were specified -- learned (L), not learned (N), and unknown (U). The way in which skills were classified and then used to predict success or failure in subsequent tutor problem sets and posttest problems was as follows. First, we examined individuals' tutor performance data (accuracy) for any instance(s) of problem-set mastery. For example, the earliest occurrence of a score of 100% on a problem set would indicate that the individual *most likely* knew the associated skills. The subject would thereby receive a set of tentative L's for those skills required by that problem set (pass 1). If the following problem set required the same skills, and the subject again demonstrated 100%, then the shared skills would all be assigned L's (pass 2). Working forward, all subsequent problem sets containing these skills would also receive designations of L. A skill received a "not learned" (N) classification if, for example, two problem sets contained all but one skill in common. If the learner's score on the first problem set was 100%, but was less than 100% on the second problem set which contained the new skill, that skill would be classified as N. Subjects' data that did not have 100% on *any* of the 14 problem sets were problematic (but not impossible) to model. These individuals' data (note: only 30 out of 370 subjects did not have 100% on any problem set) would start out with a lot of unknown (U) status skills. Beginning with the highest obtained accuracy and working backwards and forwards (i.e., deductive and inductive reasoning) through the problem sets, we could begin to make L and N assignments to those unknown skills, as well. We now illustrate the modeling procedure, predicting within-tutor, then posttest performance data.

Modeling Within-tutor Performance. We randomly selected a subject's data. "Random Joe" (not his real name)[2] performed at 100% accuracy on the first five problem sets of the tutor (see Table 2, above). Consider his performance data for the first problem-set cluster (Scale Conversion problems: G-1, G-2, G-3). Because the first 100% occurred at G-1, and he also acquired skills from the second problem set at 100% accuracy, we designated all four shared skills as L. Regarding intra-cluster predictability, G-2 required the exact same skills as G-1, but had a stricter accuracy criterion. Our prediction was that G-2 should be slightly less than (or equal to) 100%. In fact, Joe's actual G-2 score was 100%, showing that he could handle the more stringent criterion. Next, G-3 required the same four skills as in G-1 and G-2. However, G-3 represented an *analog* problem to G-1 and G-2 (i.e., the skills were the same, but the subject matter was different). The subject matter introduced in G-3 involved altimeter settings, so this problem set had an additional skill requirement (Skill 5 = U). We predicted that Joe would score slightly less than (or equal to) 100% on G-3 because (a) it was an analog problem with one additional skill requirement making it harder than G-1 and G-2, and (b) he had already demonstrated mastery on four (out of five) required skills. It turns out that Joe, again, performed at 100% accuracy on this problem set. Because he had successfully acquired skills 1 - 4 (and tentatively acquired skill 5), we designated those skills as L in all subsequent problem sets including them.[3]

The same analysis was employed on problem sets G-4, G-5, G-6, and G-14 representing the Cartesian Coordinate Grid cluster. Joe scored 100% on problem sets G-4 and G-5, so we designated the shared skills (6, 7, 8) as L within these two problem sets, as well as

---

[2] *Random Joe* had been assigned to the Abbreviated condition, and was randomly selected out of 364 cases.

[3] Skill 5 was not required for use again until problem set G-6. Because skill 5 had received a "Tentative L" assignment, and Joe's accuracy score for G-6 was 14%, the tentative status for skill 5 remained.

in all subsequent problem sets. Skill 9 (Plot point at intersection) was a partially-shared skill. That is, it was first required in problem set G-4, and because Joe was 100% at G-4, skill 9 received a tentative L. However, because skill 9 wasn't needed in G-5, and in G-6 (where it was next needed), Joe did not perform very well, skill 9's tentative status remained (note: it wasn't until G-8 that the status would change to a confirmed L). Problem set G-6 represented an analog problem set, using the same subject matter as was introduced in G-3, but with a lot of new skill requirements. That is, G-6 not only employed skills used in G-4 and G-5, but also introduced three new (and difficult) skills (11, 12, 13). Therefore, we predicted that Joe's accuracy score would be considerably less than 100%. Joe's obtained score of 14% was in line with the prediction, attributed to the difficulty level of these three new skills that dealt with the manipulation of the correction table and separation of the altimeter setting into component parts. The new skills were designated N. The final problem set in this cluster, G-14, utilized the same basic skills as the others (G-4, G-5, G-6), but was situated in a completely new context -- the Maximum Allowable Crosswind chart. All of G-14's requisite skills had received L's (by forward propagation) except for one new skill (25 = U). So, given that there was just one new skill introduced in G-14, and it represented an analog to earlier problem sets, we predicted that Joe would score slightly less than 100% on G-14. In fact, Joe scored 75%, so skill 25 was designated N. While there were no other problem sets in the tutor requiring skill 25, the posttest contained six items that did, so this skill status should be predictive later on.

The remaining problem sets were similarly analyzed. In summary, the skills classified for Joe as learned (L) were: 1, 2, 3, 4, 6, 7, 8, 9, 10, 13, 14, 15, 17, and 18. The only skill classified as tentatively learned (L?) was skill 5. Skills classified as not learned (N) included skills: 11, 12, 15, 22, 23, 24, and 25. Finally, skills that had an unknown (U) status (from a lack of information) were: 19, 20, and 21. This represented Joe's knowledge state at the end of the Graph section of the tutor. However, he still had to complete the TOLD section of the tutor before taking the posttest, and that would give him *additional* practice on these component skills. Thus, skill status changes could transpire during TOLD learning, where N's and U's could change to L's or N's. Because we consider only the Graph data in this paper, our predictions will be less precise than if the TOLD data were also included in the equation.

Modeling Posttest Performance. In order to predict posttest performance from within-tutor performance data, we first needed to link the 25 unique skills, outlined in Table 2, to each posttest item. That mapping resulted in a binary matrix consisting of 46 posttest items by 25 skills (similar to Table 2, only larger, and using posttest, rather than problem set, data).

To illustrate, posttest item 1 required a single skill -- reading data from a straight-line graph (skill 10). We had assigned an L status to Joe's skill 10, so we predicted that he would get this item correct, and he did. On the other hand, Joe had performed poorly on problem set G-6 (14%). Thus, we predicted that he would fail on any posttest item that required skills classified as "not learned" (N) from that problem set (i.e., skills 11 and 12). Our prediction was supported: Joe incorrectly solved both posttest items that involved the unlearned skills. One of these items asked:

IF YOU WANT TO KNOW THE PRESSURE ALTITUDE AND ARE GIVEN THE STATION PRESSURE (ALTIMETER SETTING) AS 850 MILLIBARS, THE NEXT STEP IN THE PROCEDURE IS TO:

(a) Find correction value to pressure altitude
(b) Convert millibars to inches of Mercury
(c) Add correction value to field elevation
(d) Convert field elevation to feet
(e) Subtract correction value from field elevation
(f) Add station pressure to correction value

The correct answer for this item was (b), but Joe incorrectly selected (e), clearly demonstrating that he confused the sequence of steps in this particular procedure. Next, we tested another skill which had been designated as L -- Skill 14 (RWD). We examined all five posttest items that involved skill 14. Joe answered 4 of the 5 relevant items correctly, showing that he really seemed to comprehend "relative wind direction." The one item that he failed to answer correctly involved not only RWD, but also Wind Type (skill 15), a skill he had <u>not</u> learned (i.e., classified as N). This confounded item asked:

IN DETERMINING WIND TYPE, YOU MUST CALCULATE THE RELATIVE WIND DIRECTION. YOU ARE GIVEN THE WIND DIRECTION, BUT ALSO NEED THE:

(a) Polar coordinate chart	(d) Runway heading
(b) Wind velocity	(e) True altitude
(c) Gust increment	(f) Altimeter setting

Joe selected response (b) indicating that he misinterpreted the question to be addressing wind component issues. That is, for wind components, you need to know the relative wind direction as well as wind velocity, which is the answer that he erroneously chose. The correct answer to this item, however, is (d) runway heading. Thus, we were correct in designating skill 14 as L and skill 15 as N.

Finally, we analyzed the ambiguous-status skills classified as U (e.g., skill 20 -- Normalizing RWD) to see whether these skills were really L or N (remember, the "U" status simply means unknown). Skill 20 was needed in the solution of three posttest items. Joe answered all of those items correctly, so we can infer that he <u>did</u> understand when and how to normalize the RWD. The probable reason this skill was assigned a U status was that Joe had shown difficulties solving problems requiring this skill during the Graph section of the tutor, seen in his less-than-perfect accuracy data on relevant problem sets (G-11 through G-13). However, this skill was also exercised during the TOLD section of the tutor, so he probably "learned" skill 20 at that time.

We are currently in the process of assigning differential weights for skills in terms of their relative importance, within and across problem sets. This weighting scheme will allow us to predict, probabilistically and more precisely, whether a subject will respond to a given problem (i.e., tutor or posttest) correctly or not based on: (a) whether the person unequivocally knew the relevant skills or not, and (b) the weighted importance of each skill to that item. For items requiring just one skill, if the skill was designated L, then the prediction for the item would simply be "correct." If the skill status was N, then we would predict "incorrect," and if U, then it could go either way. For other items containing more than one skill, we need to employ a more complex decision rule including the weights in the equation.

Discussion

In summary, we examined relations among practice, performance, and learning outcome using an intelligent tutoring system instructing flight engineering knowledge and skills. We created two contrasting practice conditions from the tutor: one extended (fairly typical), and one greatly shortened. While some of the findings were expected (e.g., the Abbreviated group required significantly less time to complete the tutor compared to the Extended group), other findings were completely unexpected (there were no significant differences between the two groups on any of the outcome measures -- accuracy or latency). In an attempt to account for

the unexpected findings, we first established (through item analysis, correlations, and so on) that the posttest was a reliable measure of acquired flight engineering knowledge and skills.

Next, we looked at practice effects on within-tutor performance data to see if there were any significant group differences that may have obscured (i.e., attenuated) outcome differences. Schmidt and Bjork (1992) have argued that differences between groups that show up during the acquisition phase may reflect either differences in learning or performance (or both). Furthermore, variations on the practice schedule (e.g., random, reduced, massed, or variable practice) often degrade performance during practice, relative to more ideal conditions. This was demonstrated in our study: wherever significant differences between groups emerged, subjects in the Abbreviated condition performed worse than subjects in the Extended condition. While the Abbreviated group did perform worse during acquisition (on the harder problem sets), their data suggest that they may have compensated for having so few problems by devoting more time per problem. Furthermore, they were actively engaged in the learning process, transferring new knowledge and skills across successive problem sets to the same degree as the Extended group. Finally, both groups' learning curves were quite predictable. For example, if a new and difficult skill was introduced in some problem set, then subjects' learning latencies and errors incremented by the same relative degree, per group. Or, if a problem set required subjects to apply some previously learned skill, latencies and errors would comparably decrement.

Our simple modeling procedure described in this paper provides a formalism for predicting transfer. In general, vertical transfer is expected to occur where subjects have demonstrated they know certain skills. For example, we predicted that Joe would acquire problem set 2 fairly easily given that (a) this problem set involved skills 1 - 4, and (b) he had demonstrated mastery of these requisite skills. Our prediction was upheld. But a person's skill designated as "learned" does not guarantee that the individual will always apply it correctly in the solution of novel problems. Nor does a skill designated as "not learned" guarantee that the person will answer relevant problems involving that skill incorrectly. The important point is that if we can predict transfer, then we may be approaching an understanding of transfer mechanisms. An understanding of transfer mechanisms will allow us to capitalize on the mechanisms in operational settings, such as for: (1) predicting transfer among jobs by capitalizing on a person's incoming knowledge and skills; and (2) optimizing re-training programs.

These findings suggest several important implications for the design of automated instruction, particularly for the teaching of complex skills where practice opportunities are important. Traditionally, it has been accepted that practice makes perfect, and more practice is better than less practice. In contrast, these results clearly show that in certain cases, with regard to practice opportunities, enough is enough. By adopting a policy of iterative pilot testing, developers can avoid unnecessary time and effort in providing too many practice opportunities. Such a policy would also avoid investing too much of the students' time in accomplishing unnecessary practice opportunities. Undue tedium during training almost certainly has other negative consequences as well (e.g., generalized loss of motivation, reduced time for other training needs, increased wash-out rate).

Among the more novel implications of these findings is the possibility that students of automated instruction may actually self-regulate their practice behavior to some degree. In this study, students who received fewer practice opportunities spent more time per problem than students given greater practice opportunities. It is unclear whether this represents a voluntary decision to allocate additional resources by the Abbreviated group, a function of

their making more errors, or a little of both. In any event, this result supports conventional wisdom among trainers that motivated students will learn regardless of how deficient the training is. It does not, however, negate the fact that it is possible to optimize instruction.

We believe that the most important implications of this research derive from the componential skill analysis. In general, we believe it is expedient to design instruction that focuses on the exact component skills that make up a desired outcome performance (although we will take up an important exception to this rule momentarily). Rather than simply giving students increasing amounts of practice during training, componential skill analysis allows designers to systematically design practice opportunities that hierarchically and incrementally add to the student's repertoire of critical skill components. It is then possible to monitor the acquisition of these skills during automated instruction, and then control the run-time generation of practice opportunities by simple algorithms. Carefully designed systems can deliver *just enough,* and the right kind of, practice opportunities to produce the desired outcome performance, while minimizing training time. There is, however, a class of exceptions to the rule of only teaching desired outcome skills. Sometimes it is appropriate to provide tools that ease cognitive load during certain stages of training, even though these tools will not be available in the applied setting. For instance, mastery of high-resource component skills can sometimes be facilitated by providing tools to learners that reduce the cognitive load associated with other, currently unmastered component skills. In these cases, it has been shown that temporary "crutches" during training are very beneficial (Lintern & Gopher, 1978). However, designers need to consider the added load of learning tool-specific skills, as well as the fact that the student will need to be carefully weaned from such tools to assure that the desired transfer environment (performance requirements in the applied setting) matches the terminal practice environment (the end-of-training performance requirements).

Early in this paper, we cited a common finding, "When the number or variety of example problems is restricted, skill acquisition tends to be rapid, but transfer tends to be weak." In our data, we documented the rapidity, but not the weakness. Abbreviated subjects finished the tutor significantly faster than the Extended subjects, but the two groups attained the same degree of transfer performance for both vertical (within-tutor) and lateral (outcome) transfer. In conclusion, these data suggest that for even this complex material, if the curriculum is developed in a hierarchical manner, based on (and requiring mastery of) hierarchical skills which have been decomposed via a careful task analysis, even a minimum number of practice opportunities may suffice.

REFERENCES

Ackerman, P.L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. ***Journal of Experimental Psychology: General, 117***, 288-318.

Anderson, J.R. (1987). Skill acquisition: Compilation of weak-method problem solutions. ***Psychological Review, 94***, 192-210.

Bryan, W.L., & Harter, N. (1899). Studies on the telegraphic language: The acquisition of a hierarchy of habits. ***Psychological Review, 6***, 345-375.

Carlson, R.A., & Yaure, R.G. (1990). Practice schedules and the use of component skills in problem solving. ***Journal of Experimental Psychology: Learning, Memory, & Cognition, 16***, 484-496.

Fisk, A.D. & Rogers, W.A. (1992). The application of consistency principles for the assessment of skill development. In J.W. Regian & V.J. Shute (Eds.), ***Cognitive approaches to automated instruction*** (. pp. 171-194). Hillsdale, NJ: Lawrence Erlbaum.

Gagné, R.M., & Paradise, N.E. (1961). Abilities and learning sets in knowledge acquisition. ***Psychological Monograph, 75***(14).

Gick, M.L. & Holyoak, K.J. (1987). The cognitive basis of knowledge transfer. In S.M. Cormier & J.D. Hagman (Eds.), ***Transfer of Learning***. New York: Academic Press.

Lesgold, A., Bunzo, M.S., McGinnis, T. & Eastman, R. (1989). ***Windy, the Flight Engineering Tutor*** [Unpublished tutor]. Pittsburgh, PA: Learning Research and

Development Center, University of Pittsburgh.

Lintern, G. and Gopher, D. (1978). Adaptive training of perceptual motor skills: Issues, results and future directions. ***International Journal of Man-Machine Studies, 10***, 521-551.

Schmidt, R.A. and Bjork, R.A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. ***Psychological Science***, 3(4), 207-217.

Schneider, W., & Shiffrin, R.M. (1977). Controlled and automatic human information processing. Detection, search, and attention. ***Psychological Review, 84***, 1-66.

Shute, V.J. (1993). A macroadaptive approach to tutoring. ***Journal of Artificial Intelligence & Education, 4***(1), 61-93.

Singley, M.K. & Anderson, J.R. 1989. ***The transfer of cognitive skill***. Cambridge, MA: Harvard Press.

Smith, S. (1986). ***An analysis of transfer between tower of Hanoi isomorphs***. Unpublished doctoral dissertation, Carnegie-Mellon University.

Woltz, D.J. (1988). An investigation of the role of working memory in procedural skill acquisition. ***Journal of Experimental Psychology: General, 117***, 319-331.

Appendix 1: Listing of Skills for all Problem Sets of the Tutor -- Graph Section

Scale conversion Problems:

G-1 (Converting C° to F° and F° to C°), Accuracy: ±2 degrees
Skill 1: Find given value on conversion scale; Skill 2: Read opposite side of scale for equivalent value (temp); Skill 3: Interpolate/determine value; Skill 4: Input (converted) value.

G-2 (Converting C° to F° and F° to C°), Accuracy: ±1 degree
Skill 1: Find given value on conversion scale; Skill 2: Read to opposite side of scale for equivalent value (temp); Skill 3: Interpolate/determine value; Skill 4: Input (converted) value.

G-3 (Converting Millibars to In. Hg), ANALOG PROBLEM (to G-1, G-2), Accuracy: ±1 in. Hg
Skill 5: Concept: altimeter setting; Skill 1: Find given value on conversion scale; Skill 2: Read to opposite side of scale for equivalent value (In. Hg); Skill 3: Interpolate/determine value; Skill 4: Input (converted) value;

Using Cartesian Coordinate Grid:

G-4 (Plotting x, y coordinates), Accuracy: ±1 degree
Skill 6: Cartesian Coordinate chart; Skill 7: Use vertical ruler (to draw line on x-axis from given x coordinate); Skill 8: Use horizontal ruler (to draw line on y-axis from given y coordinate); Skill 3: Interpolate/determine value (between marked values of x- and y-axis); Skill 9: Plot point at intersection (of horiz. and vert. lines).

G-5 (Using straight-line graph to convert C° and F° temperatures), Accuracy: ±1 degree
Skill 6: Cartesian Coordinate chart; Skill 10: Concept: Straight-line graph (linear relation between scales); Skill 7 (or 8): Use vertical/horizontal ruler (to intersect straight-line at given x- or y-coordinate); Skill 8 (or 7): Use opposite ruler (to draw line from straight-line to axis); Skill 3: Interpolate/determine value (from intersection of line with axis); Skill 4: Input (converted) value.

G-6 (Plotting x, y coordinates), ANALOG PROBLEM (to G-4, G-5), Complete accuracy
Skill 5: Concept: altimeter setting; Skill 11: Correction table (correcting pressure altitude); Skill 12: Separate altimeter setting into x & y coordinates (1st 3 digits = y-axis; 4th digit = x-axis); Skill 7: Use vertical ruler (button on x- axis to highlight column); Skill 8: Use horizontal ruler (button on y- axis to highlight row); Skill 9: Plot point at intersection (of column and row); Skill 4: Input (correction) value; Skill 13: Compute value (pressure altitude = field elevation + correction value); Skill 4: Input (computed) value.

G-14 (Determining maximum allowable XW), Accuracy: ±1 knot
Skill 6: Cartesian Coordinate Grid; Skill 10: Straight-line graph; Skill 7: Use vertical ruler (to intersect straight-line at given x coordinate/GW of aircraft); Skill 25: Locate given RCR on chart; Skill 8: Use horizontal ruler (to draw line from straight-line to axis); Skill 3: Interpolate/determine value (where horizontal line intersects y-axis/Max Allowable XW; Skill 4: Input (Max XW) value.

Using Polar Coordinate Chart:

G-7 (Computing Relative Wind Direction -- RWD, determining wind type), Complete accuracy
Skill 14: Concept: RWD; Skill 13: Compute value (RWD = |wind direction-runway heading|); Skill 4: Input (computed) value; Skill 15: Determine wind type (according to rules); Skill 4: Input (wind type) value.

G-8 (Plotting x, y coord's), Accuracy: ±1 degree
Skill 16: Polar Coordinate Chart; Skill 3: Interpolate/determine value (corresponding to given RWD value); Skill 17: Plot vector (corresponding to given RWD value); Skill 18: Locate arc on PCC; Skill 9: Plot point at intersection (of RWD angle & wind velocity arc).

G-9 (Computing RWD, plotting x, y coord's), Accuracy: ±1 degree
Skill 16: Polar Coordinate Chart; Skill 14: Concept: RWD; Skill 13: Compute value (RWD); Skill 4: Input (computed) value; Skill 3: Interpolate/determine value (corresponding to computed RWD value); Skill 17: Plot vector (corresponding to computed RWD value); Skill 18: Locate arc on PCC; Skill 9: Plot point at intersection (of RWD angle & wind velocity arc).

G-10 (Normalizing RWD, plotting coord's on quartered PCC), Accuracy: ±1 degree
Skill 16: Polar Coordinate Chart; Skill 19: Quartered symmetrical PCC (folded in 1/2 and 1/4); Skill 20: Concept: normalizing RWD; Skill 13: Compute value (normalized RWD); Skill 4: Input (computed) value; Skill 3: Interpolate/determine value (RWD vector); Skill 21: Draw new vector on quartered PCC (between marked rays); Skill 17: Plot vector (corresponding to normalized RWD value); Skill 18: Locate arc on PCC; Skill 9: Plot point at intersection (of RWD angle & wind velocity arc).

Using Wind Components Chart:
G-11 (Determining HW component), Accuracy: ±1 knot
Skill 22: Wind Component Chart (PCC + Cartesian); Skill 6: Cartesian Coordinate chart; Skill 19: Quartered PCC (folded in 1/2 and 1/4); Skill 23: Gust Element (Add it to wind velocity for TW's & XW's, but not HW's); Skill 14: Concept: RWD; Skill 13: Compute value (RWD); Skill 4: Input (computed) value; Skill 20: Concept: normalizing RWD; Skill 13: Compute value (normalized RWD); Skill 4: Input (computed) value; Skill 3: Interpolate/determine value (RWD vector); Skill 21: Draw new vector on PCC (between marked rays); Skill 17: Plot vector (corresponding to normalized RWD value); Skill 18: Locate arc on PCC; Skill 9: Plot point at intersection (of RWD angle & wind velocity arc); Skill 24: Extraction location (HW & TW comp.'s from y-axis, XW comp. from x-axis); Skill 8: Use horizontal ruler (to draw line from vector to y-axis); Skill 3: Interpolate/determine (HW component) value; Skill 4: Input (HW component) value.

G-12 (Determining XW component), Accuracy: ±1 knot
Skill 22: Wind Component Chart (PCC + Cartesian); Skill 6: Cartesian Coordinate chart; Skill 19: Quartered PCC (folded in 1/2 and 1/4); Skill 14: Concept: RWD; Skill 13: Compute value (RWD); Skill 4: Input (computed) value; Skill 20: Concept: normalizing RWD; Skill 13: Compute value (normalized RWD); Skill 4: Input (computed) value; Skill 3: Interpolate/determine value (RWD vector); Skill 21: Draw new vector on PCC (between marked rays); Skill 17: Plot vector (corresponding to normalized RWD value); Skill 23: Gust Element (Add it to wind velocity for TW's & XW's, but not HW's); Skill 13: Compute value (wind velocity + gust increment); Skill 18: Locate arc on PCC; Skill 9: Plot point at intersection (of RWD angle & wind velocity arc); Skill 24: Extract XW comp. from x-axis; Skill 7: Use vertical ruler (to draw line from vector to axis); Skill 3: Interpolate/determine (XW component) value; Skill 4: Input (XW component) value.

G-13 (Determining TW component), Accuracy: ±1 knot
Skill 22: Wind Component Chart (PCC + Cartesian); Skill 6: Cartesian Coordinate chart; Skill 19: Quartered PCC (folded in 1/2 and 1/4); Skill 14: Concept: RWD; Skill 13: Compute value (RWD); Skill 4: Input (computed) value; Skill 20: Concept: normalizing RWD; Skill 13: Compute value (normalized RWD); Skill 4: Input (computed) value; Skill 3: Interpolate/determine value (RWD vector); Skill 21: Draw new vector on PCC (between marked rays); Skill 17: Plot vector (corresponding to normalized RWD value); Skill 23: Gust Element (Add it to wind velocity for TW's & XW's, but not HW's); Skill 13: Compute value (wind velocity + gust increment); Skill 18: Locate arc on PCC; Skill 9: Plot point at intersection (of RWD angle & wind velocity arc); Skill 24: Extract TW comp. from y-axis; Skill 8: Use horizontal ruler (to draw line from vector to y-axis); Skill 3: Interpolate/determine (XW component) value; Skill 4: Input (TW component) value.